

Spatial distribution of aboveground biomass stock in tropical dry forest in Brazil

Héveli Kalini Viana Santos⁽¹⁾,
Robson Borges de Lima⁽²⁾,
Rafael L Figueiredo de Souza⁽³⁾,
Domingos Cardoso⁽⁴⁾,
Peter W Moonlight⁽⁵⁾,
Thaine Teixeira Silva⁽¹⁾,
Cinthia Pereira de Oliveira⁽²⁾,
Francisco Tarcísio Alves Júnior⁽²⁾,
Elmar Veenendaal⁽⁶⁾,
Luciano Paganucci de Queiroz⁽⁷⁾,
Priscyla MS Rodrigues⁽⁸⁾,
Rubens Manoel dos Santos⁽⁹⁾,
Tiina Sarkinen⁽¹⁰⁾,
Alessandro de Paula⁽¹⁾,
Patrícia Anjos Bittencourt Barreto-Garcia⁽¹⁾,
Toby Pennington⁽¹¹⁾,
Oliver L Phillips⁽¹²⁾

Climate change is being intensified by anthropogenic emission of greenhouse gasses, highlighting the value of forests for carbon dioxide storing carbon in their biomass. Seasonally dry tropical forests are a neglected, threatened, but potentially critical biome for helping mitigate climate change. In South America, knowing the amount and distribution of carbon in Caatinga seasonally dry vegetation is essential to understand its contribution to the global carbon cycle and subsequently design a strategic plan for its conservation. The present study aimed to model and map the spatial distribution of the potential forest biomass stock across 32 forest fragments of Caatinga, in the state of Bahia, Brazil, using regression kriging and Inverse Square of Distance techniques, building from point measurements of vegetation biomass made on-the-ground in ecological plots. First, a model for estimating biomass was fitted as a function of environmental variables to apply regression kriging, and then applied to the maps of the selected components. Elevation, temperature, and precipitation explained 46% of the biomass variations in the Caatinga. The model residuals showed strong spatial dependence and were mapped based on geostatistical criteria, selecting the spherical semivariogram model for interpolation by ordinary kriging. Biomass was also mapped by the Inverse Square of Distance approach. The quality of the regression model suggests that there is good potential for estimating biomass here from environmental variables. The regression kriging showed greater detail in the spatial distribution and revealed a spatial trend of increasing biomass from the north to south of the domain. Additional studies with greater sampling intensity and the use of other explanatory variables are suggested to improve the model, as well as to maximize the technique's ability to capture the actual biomass behavior in this newly studied seasonally dry ecosystem.

Keywords: Geostatistics, Regression Kriging, Spatial Analysis, Forest Inventory

□ (1) Universidade Estadual do Sudoeste da Bahia, Departamento de Ciências Florestais, 45031-900 Vitória da Conquista, BA (Brazil); (2) Universidade do Estado do Amapá, Departamento de Engenharia Florestal, 68900-070 Macapá, AP (Brazil); (3) Universidade de São Paulo, Escola Superior de Agricultura Luiz de Queiroz, 13418-900, Piracicaba, SP (Brazil); (4) Instituto de Biologia, Universidade Federal da Bahia, 40.026-010, Salvador, BA (Brazil); (5) Tropical Diversity Section, Royal Botanic Garden Edinburgh, EH3 5NZ, Edinburgh (United Kingdom); (6) Wageningen University, Plant Ecology and Nature Conservation Group, 6700 AK, Wageningen (Netherlands); (7) Departamento de Ciências Biológicas, Universidade Estadual de Feira de Santana, 440369-00, Feira de Santana, BA (Brazil); (8) Colegiado de Ecologia, Universidade Federal do Vale do São Francisco, 563049-17, Senhor do Bonfim, BA (Brazil); (9) Universidade Federal de Lavras, 37200-900, Lavras, MG (Brazil); (10) Tropical Diversity Section, Royal Botanic Garden Edinburgh, EH3 5NZ, Edinburgh (United Kingdom); (11) University of Exeter, EX4 4QG, Exeter (United Kingdom); (12) School of Geography, University of Leeds, LS2 9JT, Leeds (United Kingdom)

@ Robson Borges de Lima (robson.lima@ueap.edu.br)

Received: Mar 24, 2022 - Accepted: Feb 14, 2023

Citation: Viana Santos HK, Borges de Lima R, Figueiredo de Souza RL, Cardoso D, Moonlight PW, Teixeira Silva T, Pereira de Oliveira C, Alves Júnior FT, Veenendaal E, de Queiroz LP, Rodrigues PMS, dos Santos RM, Sarkinen T, de Paula A, Barreto-Garcia PAB, Pennington T, Phillips OL (2023). Spatial distribution of aboveground biomass stock in tropical dry forest in Brazil. *iForest* 16: 116-126. - doi: [10.3832/ifor4104-016](https://doi.org/10.3832/ifor4104-016) [online 2023-04-17]

Communicated by: Emanuele Lingua

Introduction

Seasonally dry tropical forests (SDTFs) act as important CO₂ sinks as they promote carbon fixation from the atmosphere in their plant biomass and contribute to the quantification and cycling of nutrients (Althoff et al. 2018). Knowing the biomass retained in plants allows us to understand carbon dynamics, thus providing a better understanding of the impacts of deforestation on global warming. In addition, they provide helpful information to guide forest management practices and inform decision-making in public and non-governmental sectors. For example, many efforts including floristic inventory and vegetation monitoring initiatives in tropical regions, such as the ForestPlots Plot Network (ForestPlot.net 2021) and SDTF research protocols (DRYFLOR et al. 2016, Moonlight et al. 2021), have motivated civil society and government agencies to promote sustainable environmental land management through tree planting and ecological restoration (Crouzeilles et al. 2016). Setting targets and assessing the proportionate contribution of such projects requires a solid initial understanding of the potential

stock of aboveground biomass for each ecosystem at both regional and global scales (Crowther et al. 2015).

While the importance of this information for the conservation, management, and sustainable use of SDTFs is recognized (Da Silva et al. 2017), regional spatial models of aboveground biomass (i.e., green or dry weight at tree level or species at a local scale) in exact spatial resolution are still absent in many SDTF areas, especially in the Caatinga region of northeastern Brazil. These biomass maps are needed to, in turn, create spatial understanding of variation in carbon stocks and carbon storage potential. Some maps have been developed continentally and globally (Saatchi et al. 2011), however, their relatively coarse spatial resolutions (≥ 1 km) limit their applications for local forest management, as does the fact that they contain relatively little information about the distribution potential of biomass in space (Virgens et al. 2016) and its relationships with the region's intrinsic environmental factors or gradients.

Measuring environmental variables is an essential step towards understanding on how to better quantify ecological patterns and the distribution of biomass and carbon (De Meira Junior et al. 2020). For example, positive relationships between biomass stock and environmental factors are shared across various vegetation types and ecological gradients (Slik et al. 2009). Despite encompassing the largest expanse of the South American SDTF biome and with the highest species richness and endemism (De Queiroz et al. 2017, Fernandes et al. 2020), *in situ* observations on biomass stock and environmental factors for the

Caatinga seasonally dry forests are, however, still incipient and little known (Oliveira et al. 2021).

The Caatinga vegetation occurs under extreme edaphoclimatic characteristics and is perhaps the most vulnerable Brazilian region to climate change (Ganem 2017). In addition, the Caatinga is highly threatened by inappropriate land use practices. Due to the high population density of the northeastern semi-arid region, its natural resources are widely exploited, which results in reduced production and biomass stocks, in addition to fragmenting the landscape into different physiognomies, stages of regeneration, and land uses (Althoff et al. 2018). This culminated in an intense degradation process that has advanced recently (Vieira et al. 2020).

Thus, there is a need to develop techniques to estimate and map biomass and, in turn, produce reliable information on the potential aboveground stock of forest formations and support estimates of carbon dynamics. This work examines the potential use of global climate datasets developed from satellite observations to map patterns and potential distribution of aboveground biomass in Caatinga vegetation. Unlike wetlands with high biomass driven primarily by climatic factors (e.g., temperature and precipitation) among other spatially explicit environmental factors or gradients, the Caatinga vegetation has lower biomass due to its adaptive metabolic and functional characteristics for the region, but a comparatively high potential for carbon storage (Santos et al. 2016). However, there is little local evidence to predict biomass patterns and potential distribution from environmental factors.

These regional characteristics favor the use of environmental variables to develop ecological and distribution models. The available remote sensing data provide spatially refined information on landscape and vegetation heterogeneity over the Caatinga SDTFs that can be readily incorporated into models to predict the potential distribution of biomass. These models are strictly mathematical or based on specific ecological theories (Fagua et al. 2021). A detailed discussion or review of these ecological models and theories is beyond the scope of this paper, and has already been fully addressed elsewhere (Elith et al. 2010, Hijmans & Elith 2017). Here, we are interested in modeling and mapping the Caatinga biomass distribution to understand how available resources and environmental factors condition these biomass patterns at the biome scale. We tested geostatistical methods that integrate remote sensing data and the location of geographic points from ground-based plot inventories in different Caatinga forest fragments. With this approach, we aim to model and map the potential distribution of biomass and evaluate the contribution of each environmental variable in predictive maps for the Caatinga SDTF in Bahia. Therefore, we started from the hypothesis that climatic and relief factors have a strong influence on aboveground biomass patterns in Caatinga vegetation. The paper is organized into four sections: (i) description of biomass and environmental data; (ii) description of spatial modeling and main environmental factors; (iii) assessment of potential range distributions of biomass; (iv) discussion on the potential distribution of total biomass for the characterization of suitable areas for conservation and forest management.

Material and methods

Characterization of the study area

The area covered by this study is located in the state of Bahia (Fig. 1) and comprises a seasonally dry vegetation widely known as Caatinga, the most species-rich and largest expanse of the South American SDTF biome (De Queiroz et al. 2017, Fernandes et al. 2020). The Caatinga domain occupies an area of 351,402 km² in the state, covering about 62% of the Bahian territory (IBGE 2021).

BSh climate according to the Köppen climate classification (Alvares et al. 2013) is predominant in this area, characterized as a dry and hot semi-arid climate, which associates a scarcity of rainfall with strong insolation and high temperatures. The tropical As climate is also present, with a dry summer season, and the Aw type on the west side, with a dry winter. There are also subtropical humid zones with mild summers (Cwb climate with dry winters, and Cfb without a dry season) in a small region in the center-south, where altitudes are above 1000 m a.s.l.

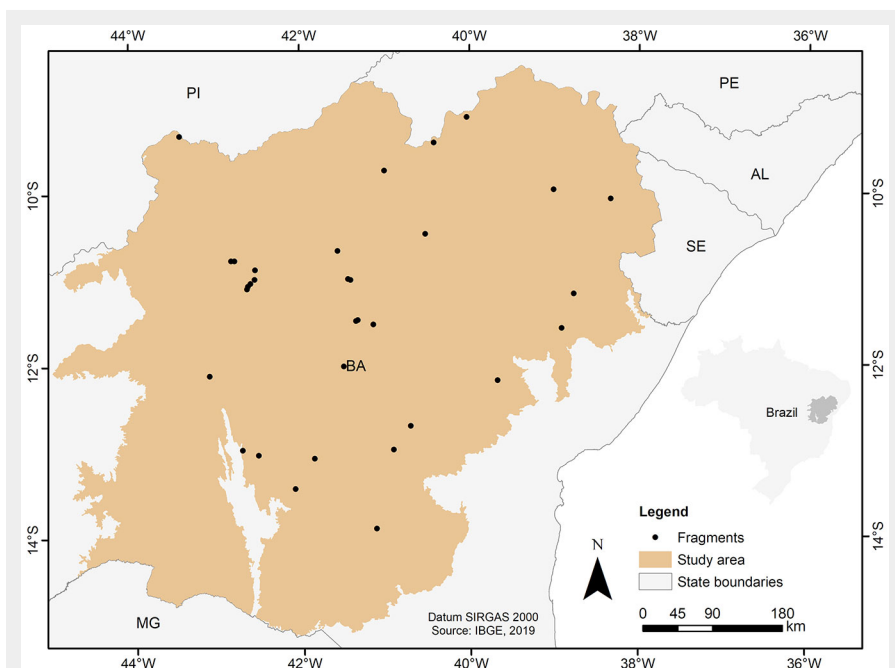


Fig. 1 - Location of the sampled fragments in the Caatinga biome, in the state of Bahia, Brazil.

There is irregularity in the rainfall distribution in time and space. This region receives from 300 to 1180 mm of annual rainfall; however, an expressive part of the area receives average rainfall of less than 750 mm year⁻¹. Moreover, the precipitation is less than 400 mm per year in a small part. About 50% of these annual totals are generally concentrated in just three months (Fick & Hijmans 2017).

Temperatures also vary in space, with annual averages between 16 and 29 °C, however, they are predominantly above 24 °C (Fick & Hijmans 2017). There are plateaus and mountains which can exceed 1000 m in altitude. Among these geomorphologies are depressions which occur more expressively throughout the area. The relief has an influence on the spatial patterns of temperatures (Castanho et al. 2020).

The soil orders that occur are: Neosols (typical in the semi-arid environment of northeastern Brazil, characterized by low natural fertility, low water retention capacity and low levels of organic matter), Latosols, Planosols, and even small fragments of Cambisols, Argisols and Luvisols (Marques et al. 2014, IBGE 2021).

The most representative vegetation is steppe savanna, which has a deciduous, gray and thorny physiognomy in the dry season, interspersed with cactus and bromeliads. This landscape is interrupted by seasonal semi-deciduous and deciduous forest fragments which occupy regions of soils with rocky outcrops and karst relief associated with more clayey soils, and have the presence of limestone in the valleys and slopes. The latter is more expressive on the west side of the study area. There is the occurrence of savanna and small dense ombrophilous forest fragments in mountainous regions, swamps and milder climatic pockets, common in the center-south of the state of Bahia (IBGE 2021).

Biomass data estimates

The database for biomass estimation was obtained from forest surveys carried out within the study area, which include a total of 836 plots distributed in 32 forest fragments (Fig. 1). A total of 21 forest inventories carried out between 2016 and 2020 were provided by the Institute for the Environment and Water Resources of the State of Bahia (INEMA) to compose this database. The number and size of plots varied between inventories at 0.02 to 0.1 ha. Tree individuals were measured in terms of diameter at breast height (DBH), measured at 1.30 m above ground level, adopting DBH greater than or equal to 10 cm as an inclusion criterion (Fig. 2).

Data from 10 permanent plots of up to 1 ha were added to this database, registered in ForestPlots.net (Lopez-Gonzalez et al. 2011, ForestPlot.net 2021), and which contribute to the Nordeste and Dryflor projects. These plots were inventoried between 2017 and 2018. Each plot for this data was assumed to be a forest fragment

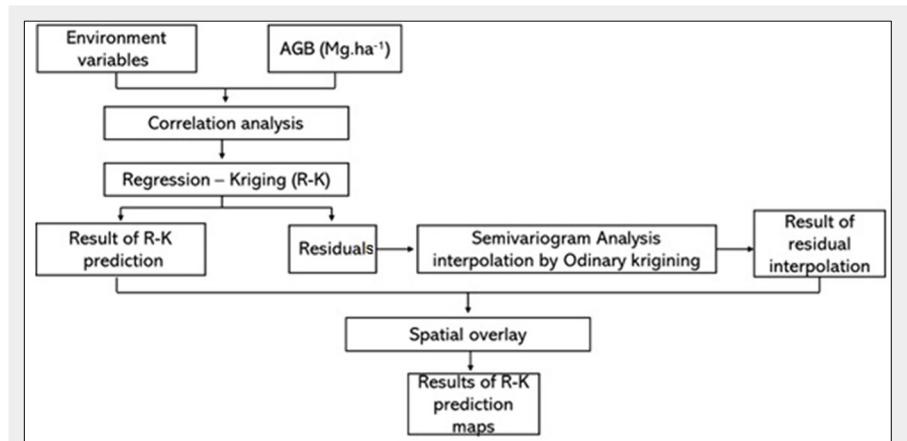


Fig. 2 - Flowchart of Regression Kriging (GWR-K) and AGB mapping predictions in this study.

and only trees with DBH ≥ 10 cm were considered.

The Contendas do Sincorá National Forest was still considered as a single forest fragment in a survey carried out in 2015, inventoried by a partnership between the Caatinga Forest Management Network and the Forestry Soil and Forest Ecology and Protection Laboratories of the University State of Southwest Bahia (UESB). The measurement of arboreal individuals with DBH ≥ 15 cm was performed for this survey.

The equation developed by Sampaio & Silva (2005) was applied to the individual data of the trees in the plots to predict the above ground woody biomass. The equa-

tion is specific to the Caatinga vegetation and is the most used by the scientific community (eqn. 1):

$$AGB = 0.1730 \cdot DBH^{2.295} \quad (1)$$

(R² = 92%, RMSE = 3.8 kg, bias = 7.6%) in which AGB is the aboveground dry biomass for each tree individual (kg), DBH is the diameter measured at 1.30 m above the ground (cm) and R² is the coefficient of determination. RMSE (root mean squared error) is the measure that calculates the “root mean squared error” of the errors between observed values (actual) and predictions (hypotheses).

Tab. 1 - Bioclimatic variables and elevation above sea used to adjust the biomass prediction model in the caatinga, Bahia.

Variables	Descriptions
BIO 1	Average annual temperature (°C)
BIO 2	Average daytime range (monthly average) (°C)
BIO 3	Isothermal (Bio 2/ Bio 7) (-100) (°C)
BIO 4	Temperature seasonality (standard deviation·100) (°C)
BIO 5	Maximum temperature of the warmest month (°C)
BIO 6	Minimum temperature of the coldest month (°C)
BIO 7	Override temperature range (Bio 5 - Bio 6) (°C)
BIO 8	Average temperature of the wettest quarter (°C)
BIO 9	Average temperature of the driest quarter (°C)
BIO 10	Average temperature of the warmest quarter (°C)
BIO 11	Average temperature of the coldest quarter (°C)
BIO 12	Average annual rainfall (mm)
BIO 13	Rainfall in the wettest month (mm)
BIO 14	Rainfall of the driest month (mm)
BIO 15	Seasonality of precipitation (Coefficient of variation) (mm)
BIO 16	Precipitation of the wettest quarter (mm)
BIO 17	Rainfall in the driest quarter (mm)
BIO 18	Precipitation of the warmest quarter (mm)
BIO 19	Rainfall of the coldest quarter (mm)
Elev	Elevation (m)

The biomass stocks of trees per plot were summed and were extrapolated in ton per hectare (Mg ha^{-1}) according to the size of the plots. Due to the heterogeneity in the biomass stock of the plots within the same fragment, the dataset used in the analyzes was the average of the biomass stock of the plots in each of the 32 forest fragments (Scolforo et al. 2016).

Spatial modeling of biomass stock

The biomass stock mapping was performed using regression kriging (RK) and Inverse Squared Distance (IQD) techniques. Kriging with regression combines two processes: elaboration of a global map which illustrates the spatial behavior of the biomass, generated from a multivariate regression model; and subsequent application of ordinary kriging on the residuals generated by the regression model.

In order to introduce specific characteristics of each micro-region of the study area and provide a continuous data base, the biomass was modeled as a function of spatial variables obtained in raster layers (Tab. 1). These variables demonstrated a potentially significant relationship with the spatial distribution of forest attributes in previous studies (Silveira et al. 2019).

After estimating the biomass stock, the data were associated with a set of predictor variables at each location. Twenty-eight geospatial covariates were selected and grouped into climatic and topographical subsets. The covariates were obtained using satellite remote sensing and globally distributed terrestrial weather stations in raster format. Each raster layer is a spatially explicit grid image, where each pixel represents the value of the described co-

variate.

The covariates were grouped into two distinct categories: topographical and climatic. Topographic covariates will include elevation, terrain slope, aspect (such as north and east), latitude (such as absolute latitude value), and a terrain roughness index (IRT). The climatic covariates will be composed of potential evapotranspiration, solar radiation, wind speed, cloud cover, and the set of 19 bioclimatic variables, 11 of which are derived from temperature: Average annual temperature, Average daytime interval (Monthly Average: max temp - min temp), Isothermal, Temperature Seasonality (standard deviation \cdot 100), Maximum temperature of the hottest month, Minimum temperature of the coldest month, Annual temperature range, Average of the wettest quarter, Average of the driest quarter, Average of the warmest quarter of the year and Average of the coldest quarter of the year; and eight variables derived from Rainfall: Annual Rainfall, Rainfall in the coldest month, Rainfall in the driest month, Seasonality of Rainfall (Coefficient of variation), Rainfall in the wettest quarter, Rainfall in the driest quarter, Rainfall in the hottest quarter and Rainfall of the coldest quarter). The complete description of the topographic covariates can be consulted at <http://earthenv.org>, and for the climatic covariates, more information can be obtained at <https://www.worldclim.org>.

All variables acquired to fit the model, except latitude and longitude, were processed in the R software (R Core Team 2021). To obtain the values of the geospatial covariates corresponding to each study site (at the plot and fragment level), they were initially obtained for the state of

Bahia, and later, using the coordinates of each site, the extraction of environmental values was performed. For each point sampled using the “raster::extract” function of the R “raster” package (Hijmans 2021) in each raster file of the geospatial covariates, this information was then stored and saved in a final matrix and used as predictor variables in the model.

The biomass values used in the modeling were submitted to exploratory analysis to identify the presence of outliers and their influence on the regression assumptions. Multiple linear regression analysis was used to adjust the model, and the parameters were estimated using the least squares method. The stepwise technique was used based on the Akaike Information Criterion (AIC) with both directions in order to define which variables best explain the biomass to construct a multivariate model.

The Shapiro-Wilk and Breusch-Pagan tests were applied to verify if the model residuals presented normality and homoscedasticity, respectively. In addition, the model was submitted to the Variance Inflation Factor (VIF) test to analyze possible correlations between the explanatory variables and ensure that the model is free of multicollinearity.

The coefficient of determination (R^2), Mean Absolute Error (MAE) and graphical analysis of residuals were evaluated to test the accuracy of the fitted model. We also used the scatter plot of the predicted values around the 1:1 line to observe the behavior of the predictions made by the model. All analyzes were performed using the R software program (R Core Team 2021).

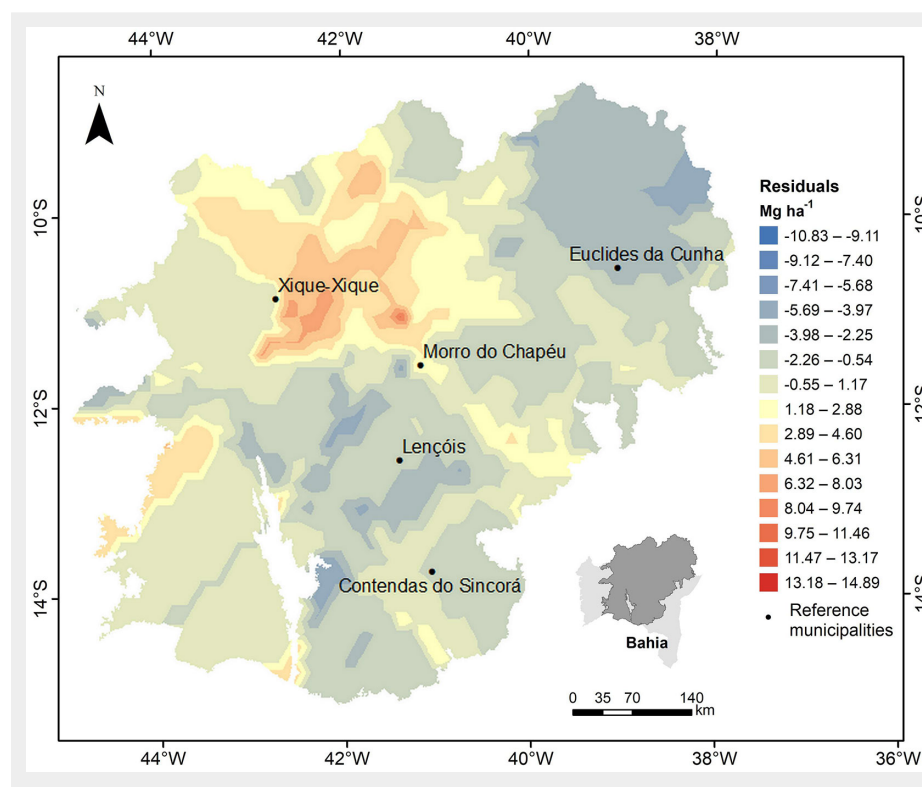


Fig. 3 - Mapping by ordinary kriging of the residuals of the regression model.

Spatialization of the biomass stock

The maps of the variables selected in the multivariate model were designed for the study area in raster format, with a spatial resolution of 1 hectare (100 × 100 m). The regression model was applied on these cells, making it possible to estimate the biomass stock in each pixel and consequently generate a global predictive map of aboveground biomass. However, this map needs to be detailed in the regression kriging procedure by adding a spatial distribution map of the regression model residuals.

Next, an exploratory analysis of the data was used to apply geostatistics on the residuals in order to know their distribution, trends and identify atypical observations (outliers). This preliminary analysis is fundamental for decision making in geostatistical procedures.

The classic estimator of Matheron (1963) was used to calculate the experimental semivariogram of the residuals, and the range was limited to half the maximum distance between two points. The behavior of semivariograms constructed in four horizontal directions was then verified to investigate the existence of anisotropy in the residuals: 0°, 45°, 90° and 135° (Guedes et al. 2013).

The spherical, exponential, and Gaussian theoretical models were directly fitted to the data using the Maximum Likelihood method, considering the stationarity assumption of the intrinsic hypothesis (Journal & Huijbregts 1978). This fitting method estimates the parameters as a function of the original data set and does not show dependence on the experimental semivariogram points; however, it requires data normality (Scolforo et al. 2016). After the fittings, the semivariogram parameters were determined and the ratio between the nugget effect and the threshold showed the degree of spatial dependence, in accordance with Cambardella et al. (1994).

The theoretical model for the interpolation was selected by evaluating the Akaike Information Criterion (AIC), the degree of spatial dependence, the reduced mean error (RME) and the standard deviation of the reduced mean error (SDRME) provided by Jackknife cross-validation (Morais et al. 2017). The closer to zero the RME and the closer to 1 the SDRME, the better the performance of the model. Then the regression residuals were interpolated from the parameters of the selected spatial model by ordinary kriging, obtaining the residual map to add a stochastic aspect to the global map of the biomass stock. This approach aims to improve the initial quality of the global map estimates, which may present trends in the estimates and not generate a sufficiently detailed spatial distribution of the biomass (Fig. 3).

The predictive kriging quality was assessed by the mean error (ME), mean standard error (MSE), and root mean squared error (RMSE), which measure the accuracy

of estimates and are provided by cross-validation by the *k*-fold method in the package “MASS” in R. Therefore, the final unbiased map of the spatial distribution of biomass was obtained for the entire Caatinga domain area in the state of Bahia by adding the residual map to the global map using map algebra tools. The semivariogram fitting was performed using the R software program (R Core Team 2021), with the help of the “geoR” package (Ribeiro Junior et al. 2020) and the maps and interpolation were prepared in the ArcGIS® software program (ESRI 2019).

The prediction of biomass using the deterministic interpolator was also performed using the Inverse Weighted Distance (IWD) with exponent 2, a method known as Inverse Squared Distance (ISD). Neighborhood parameters were defined for a minimum of 10 points and a maximum of 32 points, which corresponds to the total number of sampling points in the study area. Interpolation by ISD was performed in the ArcGIS® software program (ESRI 2019). Mapping performance was evaluated on visual and statistical criteria. The mean error (ME) and the Root Mean Square Error (RMSE) were used for the statistical evaluation. The ME measured the prediction bias and its values must be close to zero for unbiased predictions (Reis et al. 2020). Negative ME values suggest that the prediction technique is overestimating, while positive values point to an underestimation.

Results and discussion

Exploratory analysis

The estimated biomass stock values in the studied fragments ranged widely, from 2.85 to 80.88 Mg ha⁻¹. There is similarity between the mean (25.17 Mg ha⁻¹) and median values (20.33 Mg ha⁻¹), which points to an approximately symmetrical data distribution, but the superiority of the mean indicates that the highest amounts of biomass are further from the center compared to the lowest stocks. The marked difference between the minimum and maximum values resulted in a high coefficient of variation (73.7%), reflecting the heterogeneity of the study area. This variation portrays how the conditions of the sampled fragments are different. Different successional stages and anthropization degrees occur in these fragments, leading to variation in the number of individuals per hectare, in the rich-

ness and diversity of species, and in the diameters and heights of the trees.

The average biomass found was similar to that estimated by Souza et al. (2019) in a protected area in Pernambuco (28.48 Mg ha⁻¹), and lower than a fragment in Sergipe, which exhibited a contribution of 54.93 Mg ha⁻¹ (Oliveira 2016). Lima Júnior et al. (2014) estimated values between 5.93 and 60.74 Mg ha⁻¹ in Petrolina (PE), while Santos et al. (2016) observed a smaller range in Caicó (RN) of between 10 and 17.8 Mg ha⁻¹ of biomass. The discrepancy between the biomass stocks observed in the referenced works evidences the physiognomic variation that occurs along fragments of the Caatinga.

These differences in stocks between Caatinga areas are associated with precipitation, irregular rainfall distribution and the successional stage of the forest, which promote variation in the species biomass accumulation and distribution (Lima Júnior et al. 2014). Souza et al. (2019) confirmed that the formation biomass is highly variable because it results from a combination of the age of the forest fragment, rainfall and species richness. The authors found that older forests exposed to greater precipitation support a greater number of species and have twice the biomass when compared to successional and species-poor areas.

Modeling the biomass stock

We identified fragments with outliers for biomass removed from the database used for subsequent analyzes as they influenced the regression assumptions. Thus, 27 fragments were used for modeling and geostatistical analysis.

The parameters used in the regression model showed significant coefficients (Tab. 2), which implies the importance of the selected variables. The VIF test confirms the absence of multicollinearity. The generated model is parsimonious and retained three variables, which is desirable as it reduces the complexity and time spent collecting data and using the model.

The model presented a coefficient of determination (R²) equal to 46% and Mean Absolute Error (MAE) of 5.9%, which indicated that the model performed well. The MAE demonstrates the model's ability to make estimates closer to the real ones when its value tends to zero (Reis et al. 2020). The R² demonstrates how much the model is able to explain the observed data.

Tab. 2 - Coefficients estimated by the regression model and Variance Inflation Factor (VIF).

Parameter	Variable	Coefficient	p-value (<0.05)	VIF
b ₀	intercept	519.3	0.000346	-
b ₁	Elev	-0.0000594	0.000293	1.70
b ₂	Temp	1.831	0.045831	1.18
b ₃	Prec	-0.03326	0.016238	1.59

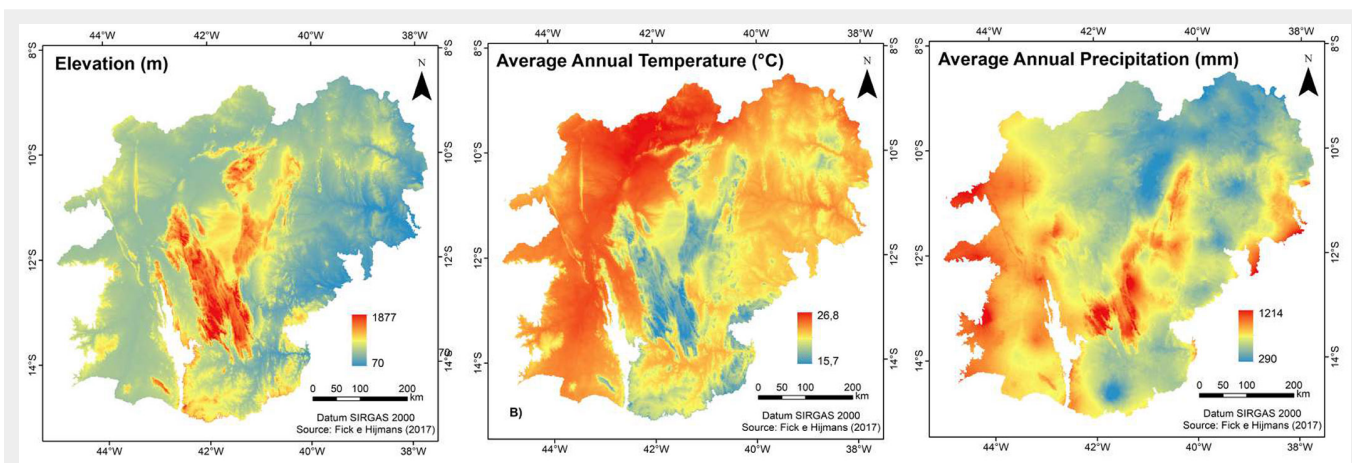


Fig. 4 - Variables selected by the multivariate regression model to estimate biomass in the Caatinga, Bahia.

The value found herein is considered acceptable given the magnitude of the study area and the wide variation that occurs between the biomass stocks of the studied fragments. Similar conditions are also reported by Scolforo et al. (2016), who found R^2 equal to 53% for a geographic model that estimates the carbon stock in the state of Minas Gerais.

The regression residuals follow a normal distribution (Shapiro-Wilk, p -value = 0.32) and are homoscedastic (Breusch-Pagan, p -value = 0.36). These characteristics are essential to define the adequacy of the model and allow adopting the Maximum Likelihood method in fitting the semivariogram, which requires *a priori* knowledge of the data distribution (Alvarenga et al. 2012). The quality of the model can also be verified in Fig. S1 (Supplementary material). The model presented randomly distributed residuals and did not show significant deviations from the mean or variance heterogeneity, suggesting that there were no underestimation or overestimation trends (Fig. S1a). These characteristics configure an estimation bias-free model. Fig. S1b illustrates the behavior of the estimated vs. the observed biomass. The closer the predicted values are to the 1:1 line, the better the fit. This means there is greater precision, less chance of estimation bias, and consequently the sum of residuals tends to zero.

Elevation, average annual temperature and average annual precipitation are the main variables responsible for the variation in the Caatinga biomass stock in the state of Bahia (Fig. 4). The contribution of precipitation is already consolidated in the literature (Barni et al. 2016, Poorter et al. 2016, Silveira et al. 2019, Souza et al. 2019). The availability of water in the soil increases the photosynthetic activity of plants, influencing the seasonality and intensity of leaf production, and consequent biomass production (Souza et al. 2019).

The prediction error associated with the developed model may be due to the preci-

sion degree of applying the allometric equation used in this study, which presents R^2 equal to 92% (Sampaio & Silva 2005). The error can also be attributed to the measuring process of the trees which was performed on different occasions and by different teams. The high variability that occurs in the biomass values, which promoted the high coefficient of variation, also represents an additional difficulty for fitting the multivariate model. This was also observed by Mello et al. (2013), who reported losses in precision statistics when modeling rainfall erosivity for the Northeast region attributed to the high variability of precipitation in this region.

Scolforo et al. (2016) and Silveira et al. (2019) proved that environmental components can be highly explanatory variables capable of estimating forest attributes. The present study corroborates these authors by identifying that there is a potential for developing models capable of estimating the biomass of the Caatinga vegetation using environmental variables. However, it is necessary to improve these models, aiming at increasing the accuracy of the estimate, given that regression equations are always associated with uncertainties (Scolforo et al. 2016). Studies using a larger sample population are suggested, which will allow construction of more robust equations with fewer errors associated with the predictive capacity.

It should be noted that the accumulation of biomass in forests does not only respond to the variables considered, but also to other environmental conditions, such as soil fertility, water deficit, seasonality of rainfall and successional stages (Poorter et al. 2016, Souza et al. 2019). The use of vegetation indices and remote sensing data is certainly a convincing option for biomass characterization and for developing regression models with lower degrees of uncertainty. Work in this direction has already been developed by Lima Júnior et al. (2014), Xue et al. (2017) who found good performance in the fitted functions.

Spatialization of the biomass stock

The semi-variance calculated in the four directions did not show significant differences and the semivariogram was considered isotropic, indicating that the spatial dependence of the regression residuals only depends on the distance between the points and is the same in all directions (Mello & Oliveira 2016).

The theoretical semivariograms fitted to the data by the Maximum Likelihood method are represented in Fig. S2 (Supplementary material). It is noteworthy that the residuals had a normal distribution, which legitimizes the use of this method (Reis et al. 2020). It is possible to detect that the residuals of the regression model are spatially structured independent from the theoretical model.

Tab. 3 shows the parameters of the fitted theoretical models and their respective degrees of spatial dependence of the residuals, in addition to the statistical criteria used to define the best model. The spherical model presented the best fit to the regression model residuals due to the lower degree of spatial dependence and nugget effect, lower AIC and reduced mean error (RME) closer to zero.

We can see the smallest nugget effect associated with the spherical model necessary to provide more accurate estimates in the kriging interpolation in Fig. S2 (Alvarenga et al. 2012). The nugget effect different from zero indicates that there is a random variation of the residual data that cannot be explained by the spatial component due to the lack of sampling in the distances smaller than the smallest distance between the plots (Scolforo et al. 2016). The lowest average distance recorded for the sampled fragments was 2447 meters.

The range found was 21,000 meters and suggests that the residuals are spatially correlated up to this distance. The high range value is due to the distances between the sampled fragments and the magnitude of the study area. The relationship between the nugget effect and the

threshold expresses the degree of spatial dependence of the variable, which is classified as strong (< 25%), moderate (25 to 75%) and weak (> 75%), according to Cambardella et al. (1994). All models showed a residual spatial dependence structure from strong to moderate, constituting desirable indices in geostatistical modeling (Mello & Oliveira 2016).

The ordinary kriging mapping of residuals by the spherical model is shown in Fig. 3. It is observed that closer points in space tend to have similar values to each other, which is the main feature revealed by geostatistics (Matheron 1963). Mello et al. (2013) highlighted that the positive values of the residuals represent underestimation of the model, while the negative values correspond to overestimation. Thus, the authors assure that a balanced distribution of these estimates is desirable, as it suggests that there are no trends produced by the regression model. A balance between under- and overestimates is noticeable on the map, showing that the model is suitable for the study area.

The error metrics associated with the interpolated surface in the ordinary kriging of residuals are presented in Tab. 4 and reinforce that the spherical theoretical model was adequate in interpolating residuals. The difference found between the RMSE and ASE indicates that the forecast was underestimated by 0.43 Mg ha⁻¹ (Mello & Oliveira 2016).

The combination of the regression model estimates (Fig. 5a) and the kriging residuals (Fig. 3) resulted in the final prediction map of the biomass stock by kriging with regression (Fig. 5b). The global map and the map corrected by the kriging of residuals reveal similar behaviors of the biomass stocks; however, the greater detail given by the final map is remarkable. The biomass stock map interpolated from the ISD technique is represented in Fig. 5c. It is characteristic of this technique to preserve the amplitude of the original data, so the minimum and maximum values found on the map are those indicated in the descriptive statistics of the data.

Tab. 3 - Statistical parameters and criteria of the theoretical models adjusted for the residuals of the regression model. (C₀): nugget effect; (C₁): threshold; (A): range; (GD): degree of spatial dependence; (EMR): mean reduced error (S_{ER}): standard deviation of reduced errors (AIC): Akaike Information Criterion.

Model	C ₀	C ₁	A (m)	GD (%)	EMR	S _{ER}	AIC
Exponential	11.75	49.00	21000	23.90	0.018	1.001	187.3
Gaussian	13.69	48.12	21000	28.45	0.020	1.002	186.3
Spherical	5.06	45.45	21000	11.14	0.016	0.992	185.4

There is generally an increase in the biomass stock in the north-south direction. Biomass stocks below 20 Mg ha⁻¹ occur in the north of the study area due to lower precipitation rates and higher temperatures. The region surrounding the municipality of Euclides da Cunha is classified by Ganem (2017) as a priority area for creating an integral protection conservation unit due to the urgency of habitat loss and susceptibility to environmental degradation.

A small area near the municipality of Xique-Xique stood out for exhibiting higher biomass values. The junction of the residual map shows the underestimation given by the global map in this region, reinforcing that the combination between a global interpolator and a geostatistician better represent the spatial distribution of a variable. The evident detail given in this specific area is due to the higher concentration of fragments sampled there, which in turn brought the regression kriging map (Fig. 5b) closer to the real stock in this region. In addition, a biomass gradient is observed in the mapping generated by kriging with regression which increases in the north-south direction. However, as this technique is dependent on the closest neighborhood points, the map did not detail the biomass for the subsampled regions and promoted the formation of islands around the sampled fragments.

The interpolated surface presented mean error (ME) values of -0.38, which suggests an overestimation of the biomass stock prediction by the ISD. The ME measures the model's tendency to under or overesti-

Tab. 4 - Statistics of the interpolation by ordinary kriging of the residuals of the regression model. (ME): Mean Error; (MSE): Mean Standardized Error; (RMSE): Root Mean Square Error Standardized; (RMSE): Root Mean Square Error; (ASE): Mean Standard Error. For the semivariogram model to be suitable for interpolation, ME and MSE must be close to 0; the RMSSE should be close to 1 e; the RMSE and the ASE should have similar values and the smallest possible (Mello & Oliveira 2016).

Statistics	Value
ME	-0.16
MSE	0.01
RMSSE	0.97
RMSE	6.39
ASE	6.82

mate a variable of interest and should be close to zero for unbiased predictions (Reis et al. 2020). The Root Mean Square Error (RMSE) was 19.83 Mg ha⁻¹. Although this type of interpolation is useful in the absence of spatial structure of the variable of interest, or even in the presence of weak dependence structure, the use of the ISD was relevant because it showed that the inconsistency in the mapping by kriging with regression was influenced by the subsampling in the study area.

Small isolated areas with smaller stocks can be observed in the south-central por-

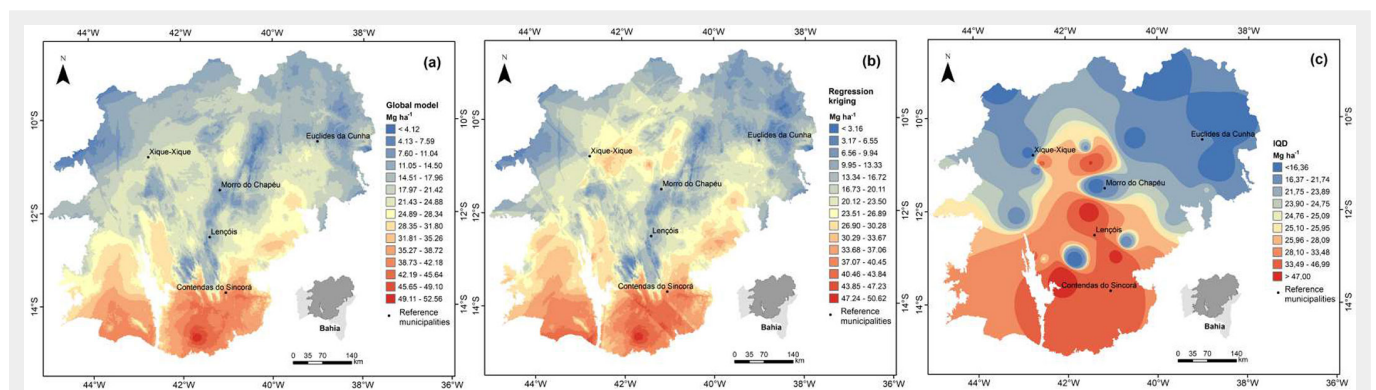


Fig. 5 - Global map of biomass stock obtained by the regression model (A), interpolated by kriging with regression - RK (B) and interpolated by the Inverse Square of Distance technique - IQD (C).

tion of the map where biomass stocks are greater than 25 Mg ha⁻¹, which constitutes an effect of the low estimated biomass in the fragments sampled there. This shows that these fragments do not capture the real variation of biomass stock that occurs in the study area and reinforces that the underestimation of the Chapada Diamantina represented in the regression kriging map is due to the effect of these fragments. Despite this, in visually comparing the maps it is notable that the regression kriging is able to capture more specific details regarding the biomass distribution in the studied area than the ISD, which makes it more efficient in the mapping.

The largest biomass stocks are in the south of the map. The biomass in the municipality of Contendas do Sincorá was estimated at 37 Mg ha⁻¹, while Virgens et al. (2016) estimated an average value of 29.2 Mg ha⁻¹. The authors classify the predominant vegetation in this area as arboreal Caatinga and in a late successional stage. Considering that these authors used the direct quantification method which is more accurate, it is stated that the biomass in the present study was slightly overestimated in this region.

The highest biomass stocks indicated in southern Bahia suggest that these are potential areas in the Caatinga to store carbon. Considering that the carbon contents in the biomass are generally around 50% (IPCC 2006), the forest carbon stock gradient in the area tends to be similar to that presented in Fig. 5b.

Low biomass values were found for the Chapada Diamantina region, where the municipalities of Morro do Chapéu and Lençóis are located. Chapada Diamantina is characterized by a set of mountains that reach more than 1000 meters in elevation, with high precipitation levels and milder temperatures (Fig. 4), constituting favorable conditions for forest productivity (Souza et al. 2019). Thus, this area tends to have larger trees, sustain denser forests and consequently store high stocks of plant biomass, as evidenced by the biomass mapping throughout the Caatinga carried out by Castanho et al. (2020) using satellite images. The authors estimated values higher than 60 Mg ha⁻¹ in this region and classified it as one of the areas with the highest plant biomass in the entire Caatinga.

However, this region does not support a homogeneity of dense vegetation. Chapada Diamantina is a set of communities which form a mosaic rich in physiognomies, depending on the topography, nature and depth of the soil. The high altitudes and busy relief result in a high erosion and occupation rate of the areas by rocky outcrops, which is limiting to vegetation development, with formations having different physiognomies from the general dominant context of this region occurring in these soils (Benites et al. 2003).

Therefore, it is likely that the low biomass

values in this area on the map (Fig. 5b) are due to fragments sampled under these conditions. A higher sampling intensity over denser forest areas would allow regression kriging to more adequately estimate the spatial behavior of the biomass stock in Chapada Diamantina, respecting its different physiognomies. Thus, the conclusions about the spatial distribution of the biomass stock in the study area are moderate, recognizing that the database is limited in the number and distribution of fragments.

However, lower mapping accuracy was expected due to the richness of physiognomies of the Caatinga in Bahia, as characterizing the biomass in heterogeneous environments such as tropical forests is a challenge (Silveira et al. 2019). But the result was satisfactory considering the database and the richness of detail achieved in the final regression kriging map.

There is no consolidated database available which covers the entire territory of the state of Bahia, which validates the data compiled in this study. Although not as consistent, the data provide the potential for biomass in each patch. A base of inventories capable of capturing the entire spectrum of vegetation variation in Bahia and showing its plurality of biomes, successional stages, anthropization degrees and phytophysiognomies is necessary. The existence of a consistent basis would enable developing techniques and studies capable of supporting the conservation and sustainable exploitation of the state's forest resources with greater precision.

This information would provide greater detail in the final map in the present study and would promote a substantial improvement in mapping. Silveira et al. (2019) and Scolforo et al. (2016) achieved satisfactory results when mapping the spatial distribution of biomass and carbon stock, respectively, by regression kriging for the state of Minas Gerais, and close to what was observed in the field. However, both authors used a robust database with greater sampling intensity and well-distributed forest fragments in the study area.

Due to the heterogeneity of the Caatinga vegetation in Bahia, it is suggested to consider the different phytophysiognomies, ecoregions or climatic zones of the biome as independent variables in the biomass modeling in future works. This distinction has been used to reduce the spatial variation common in vegetation data over large areas (Barni et al. 2016, Scolforo et al. 2016, Silveira et al. 2019). Barni et al. (2016) considered the different phytophysiognomies that occur in the state of Roraima when modeling the biomass for the state, classifying them into four forest groups that were used as input variables in the regression model. With this, the authors managed to improve the performance in the interpolation, as the model was able to better capture the vegetation variations, and thus generate a better representation of

the biomass stocks in the state.

Similarly, Scolforo et al. (2016) considered the different biomes that occur throughout the state of Minas Gerais and included them as a categorical variable in the regression model used to spatialize the carbon stock in the state. The authors obtained a consistent, more flexible model that adequately estimated the spatial behavior of carbon, with a reduction of possible trends in specific areas.

Study limitations

Our analysis explored a variety of data sources and analytical tools that can be applied to develop predictive maps that incorporate the observed spatial pattern of biomass in landscapes at the scale of the Caatinga biome in the state of Bahia. However, the context of available data still lacks more excellent coverage or prescription for some areas, therefore some limitations remain.

Firstly, it is common to face problems using inventory data collected differently in each location, mainly associated with taxonomically reliable data and correctly defined technical protocols in a given location. As a result, some trees may be measured differently, and individual biomass estimates may show under- or over-estimate trends. For example, many local, regional, and pan-tropical equations use the diameter at 1.30 m ground level (dbh) in biomass estimates, and many inventories only have a diameter at 0.30 m ground level (diameter at base). Furthermore, not all Caatinga forest fragments or land use forms can be equally represented, leading to high standard deviations or region-biased biomass estimates. Although the locations were spatially well distributed for the Caatinga biome in Bahia, and some locations were close to each other, an effort was made to correct and standardize dendrometric measurements in specialized protocols and to use forest land-use maps at a scale of finer detail. The analysis of the biomass stocks discussed is also based on accurate classifications of land cover and markedly anthropized gradients, which may not be available or may be misaligned with the inventoried plots. This helps to correct errors about the representativeness and potential distribution of biomass at the plot level and the biome scale.

Second, different scales of environmental data (10 × 10 km tables) and inventory form another complication, as plots covering a few hectares are unlikely to represent an area of 10 × 10 km (Slik et al. 2009). In addition, some regions or areas contained more inventoried trees, plots, or surveyed surface area than others, resulting in different levels of captured biomass stock by the site. Although attempts have been made to correct the diversity values for these sampling biases, likely, they have at least partially influenced the results. Ideally, meta-analyses like this should be performed on standardized plot layouts in combi-

nation with high-resolution environmental data (Gerstner et al. 2017). Unfortunately, such an approach is still complex to carry out in more remote dry tropical areas in northeastern Brazil, where reliable inventory data is already a problem (Forest-Plot.net 2021), not to mention reliable high-resolution data on soils and climate (Fick & Hijmans 2017, Hengl et al. 2017).

Thirdly, and related to the available environmental layers and their potential to explain the biomass in Caatinga vegetation, it was noted that the functional importance of many environmental variables suggests poor results among all the variables tested based on characteristics related to predictive importance. The high collinearity may help to explain this fact; however, other modeling approaches, as well as other environmental variables, should be tested, mainly related to the soil of the biome. For example, edaphic factors are vital in explaining many attributes and ecological patterns of the forest, mainly related to productivity and diversity in tropical forests (Souza et al. 2019). In addition, the low predictive values obtained for the other climatic variables may be explained by the available GIS layers in which the spatially explicit grid (interpolated values) does not correctly explain the inherent values at the plot level, requiring, therefore, an effort of data collection at local scales. Detailed investigation of these factors locally would be critical, as it could identify other ecological functions of the edaphoclimatic information set for potential biomass at a more detailed scale in the Caatinga biome of Bahia. However, the results obtained corroborate with studies of the same tropical geographic regions (Saatchi et al. 2011) with highly varied species composition (De Queiroz et al. 2017), which leads to the main conclusion being satisfactory. Finally, more direct studies on the set of ecological forces acting to define tree biomass patterns are needed to clarify these answers.

Conclusions

We integrate *in situ* tree measurements and remote sensing data to map the potential distribution of biomass in the Caatinga biome of Bahia with detailed spatial resolution. We primarily use free and open access data; therefore, our approach can be adopted and improved in other regions to assess total biomass stocks. While ecologists have an ever-expanding toolbox for extrapolating from local surveys to larger-scale patterns, it has long been assumed that the well-known barriers to reliable tree-level sampling and non-destructive estimates make sense extrapolations across the entire spectrum impossible in the Caatinga domain in the state of Bahia. Our methods enrich the information already available on biomass stocks at regional and global scales, as well as generate new perspectives for understanding and conserving tropical dry forests.

The quality of the regression model sug-

gests that three main environmental factors possibly govern aboveground biomass. Increases in temperature and altitude suggest a reduction in biomass. However, the combination of average annual precipitation and moderate altitude conditions is the ideal climate to support the highest biomass stocks. Additional studies with a larger sample population and other variables can improve the model. Compared to ISD, regression kriging revealed in more detail the spatial variability of biomass. The adoption of a higher sampling intensity has the potential to maximize this detail given by the technique. A robust base of forest inventories is needed for Bahia in order to enable studies that allow for a more accurate understanding of the potential of the state's forest resources.

Finally, we mapped the potential biomass distribution of the Caatinga biome in Bahia, but we do not guarantee that the conservation or preservation of these is established. The overwhelming effect of human activities (deforestation, fires, and the advance of agriculture and livestock) has led to substantial reductions in biomass and carbon stocks. Given the enormous carbon-storage capacity of tropical dry forests, the next avenue for research is to model the effects of climate change and future land-use changes on biomass patterns to refine conservation prospects in a changing world. These efforts are urgently needed in light of the constant and increasing rates of deforestation in the Caatinga biome in recent years and potentially threaten the planet's high biodiversity hotspots.

Acknowledgments

We thank the ForestPlots.net, Inema and Caatinga forest management network teams for permission and availability of data to conduct the research. This project was supported by ForestPlots.net approved Research Project #142: "Mapping the biomass and carbon stock, volume and diversity in tropical forests", and the UK NERC Newton Fund project "Nordeste" (NE/N012550/1). The authors thank the Co-ordination for the Improvement of Higher Education Personnel (Capes) and the National Council for Scientific and Technological Development (CNPq) for financial support in a research grant for the first and third authors of this work.

The conceptualization, data curation and formal analysis: Santos HKV, Lima RB. The Investigation, Methodology, Validation, Visualization and Project administration: Santos HKV, Lima RB, Silva TTS, Souza RLT, Oliveira CP, de Paula A, Barreto-Garcia PAB, Veenendaal E, Queiroz L, Moonlight P, Rodrigues PMS, Santos RB, Sarkinen T, Pennington T, Domingues T, Cardoso D, Phillips O. Writing-original draft: Viana HK, Lima RB. Writing review and editing: Viana HK, Lima RB, Moonlight P, Phillips O, Cardoso D.

References

- Althoff TF, Menezes RSC, Pinto AS, Pareyn FGC, Carvalho AL, Martins JCR, Carvalho EX, Silva ASA, Dutra ED, Sampaio EVSB (2018). Adaptation of the century model to simulate C and N dynamics of Caatinga dry forest before and after deforestation. *Agriculture, Ecosystems and Environment* 254: 26-34. - doi: [10.1016/j.agee.2017.11.016](https://doi.org/10.1016/j.agee.2017.11.016)
- Alvarenga LHV, Mello JM, Guedes ICL, Scolforo JRS (2012). Desempenho da estratificação em um fragmento de cerrado stricto sensu utilizando interpolador geoestatístico [Stratification performance in a cerrado stricto sensu fragment using geostatistical interpolator]. *Cerne* 18 (4): 675-681. - doi: [10.1590/S0104-77602012000400018](https://doi.org/10.1590/S0104-77602012000400018)
- Alvares CA, Stape JL, Sentelhas PC, Gonçalves JLM, Sparovek G (2013). Köppen's climate classification map for Brazil. *Meteorologische Zeitschrift* 22 (6): 711-728. - doi: [10.1127/0941-2948/2013/0507](https://doi.org/10.1127/0941-2948/2013/0507)
- Barni PE, Manzi AO, Condé TM, Barbosa RI, Fearnside PM (2016). Spatial distribution of forest biomass in Brazil's state of Roraima, northern Amazonia. *Forest Ecology and Management* 377: 170-181. - doi: [10.1016/j.foreco.2016.07.010](https://doi.org/10.1016/j.foreco.2016.07.010)
- Benites VM, Caiafa AN, Mendonça ES, Schaefer CE, Ker JC (2003). Solos e vegetação nos complexos rupestres de altitude da Mantiqueira e do Espinhaço [Soils and vegetation in the high altitude rock complexes of Mantiqueira and Espinhaço]. *Floresta e Ambiente* 10 (1): 76-85. [online] URL: <http://www.floram.org/article/588e2202e710ab87018b45f3>
- Cambardella CA, Moorman TB, Novak JM, Parkin TB, Karlen DL, Turco RF, Konopka AE (1994). Field-scale variability of soil properties in Central Iowa soils. *Soil Science Society of America Journal* 58 (5): 1501-1511. - doi: [10.2136/sssaj1994.03615995005800050033x](https://doi.org/10.2136/sssaj1994.03615995005800050033x)
- Castanho ADA, Coe M, Andrade EM, Walker W, Baccini A, Campos DA, Farina M (2020). A close look at above ground biomass of a large and heterogeneous Seasonally Dry Tropical Forest - Caatinga in North East of Brazil. *Anais da Academia Brasileira de Ciências* 92 (1): 1-18. - doi: [10.1590/0001-3765202020190282](https://doi.org/10.1590/0001-3765202020190282)
- Crouzeilles R, Curran M, Ferreira MS, Lindenmayer DB, Grelle CEV, Rey Benayas JM (2016). A global meta-analysis on the ecological drivers of forest restoration success. *Nature Communications* 7: 11666. - doi: [10.1038/ncomms11666](https://doi.org/10.1038/ncomms11666)
- Crowther TW, Glick HB, Covey KR, Bettigole C, Maynard DS, Thomas SM, Smith JR, Hintler G, Duguid MC, Amatulli G, Tuanmu M-N, Jetz W, Salas C, Stam C, Piotta D, Tavana R, Green S, Bruce G, Williams SJ, Wiser SK, Huber MO, Hengeveld GM, Nabuurs G-J, Tikhonova E, Borchardt P, Li C-F, Powrie LW, Fischer M, Hemp A, Homeier J, Cho P, Vibrans AC, Umunay PM, Piao SL, Rowe CW, Ashton MS, Crane PR, Bradford MA (2015). Mapping tree density at a global scale. *Nature* 525: 201-205. - doi: [10.1038/nature14967](https://doi.org/10.1038/nature14967)
- Da Silva JM, Leal I, Tabarelli M (2017). Caatinga: the largest tropical dry forest region in South America. Springer International Publishing, Cham, Switzerland, pp. 482. - doi: [10.1007/978-3-319-68339-3](https://doi.org/10.1007/978-3-319-68339-3)

- De Meira Junior MS, Pinto JRR, Ramos NO, Miguel EP, Gaspar RO, Phillips OL (2020). The impact of long dry periods on the aboveground biomass in a tropical forest: 20 years of monitoring. *Carbon Balance and Management*. 15: 12. - doi: [10.1186/s13021-020-00147-2](https://doi.org/10.1186/s13021-020-00147-2)
- De Queiroz LP, Cardoso D, Fernandes MF, Moro MF (2017). Diversity and evolution of flowering plants of the Caatinga domain. In: "Caatinga: The Largest Tropical Dry Forest Region in South America" (JMC da Silva, IR Leal, M Tabarelli eds). Springer International Publishing, Cham, Switzerland, pp. 23-63. - doi: [10.1007/978-3-319-68339-3_2](https://doi.org/10.1007/978-3-319-68339-3_2)
- DRYFLOR, Banda-R K, Delgado-Salinas A, Dexter KG, Linares-Palomino R, Oliveira-Filho A, Prado D, Pullan M, Quintana C, Riina R, Rodríguez MGM, Weintritt J, Acevedo-Rodríguez P, Adarve J, Alvarez E, Aranguren BA, Arteaga JC, Aymard G, Castaño A, Ceballos-Mago N, Cogollo A, Cuadros H, Delgado F, Devia W, Dueñas H, Fajardo L, Fernández A, Fernández MA, Franklin J, Freid EH, Galetti LA, Gonto R, González-MR, Graveson R, Helmer EH, Idárraga A, López R, Marcano-Vega H, Martínez OG, Maturo HM, McDonald M, McLaren K, Melo O, Mijares F, Moggi V, Molina D, Moreno Del N P, Nassar JM, Neves DM, Oakley LJ, Oatham M, Olvera-Luna AR, Pezzini FF, Dominguez OJR, Ríos ME, Rivera O, Rodríguez N, Rojas A, Särkinen T, Sánchez R, Smith M, Vargas C, Villanueva B, Pennington RT (2016). Plant diversity patterns in neotropical dry forests and their conservation implications. *Science* 353: 1383-1387. - doi: [10.1126/science.aaf5080](https://doi.org/10.1126/science.aaf5080)
- Elith J, Kearney M, Phillips S (2010). The art of modeling range-shifted species. *Methods in Ecology and Evolution* 1: 330-342. - doi: [10.1111/j.2041-210X.2010.00036.x](https://doi.org/10.1111/j.2041-210X.2010.00036.x)
- ESRI (2019). ArcGIS Desktop, version 10.8. Environmental Systems Research Institute, Redlands, CA, USA.
- Fagua JC, Jantz P, Burns P, Massey R, Buitrago JY, Saatchi S, Hakkenberg C, Goetz SJ (2021). Mapping tree diversity in the tropical forest region of Chocó-Colombia. *Environmental Research Letters*. 16: 054024. - doi: [10.1088/1748-9326/abf58a](https://doi.org/10.1088/1748-9326/abf58a)
- Fernandes MF, Cardoso D, De Queiroz LP (2020). An updated plant checklist of the Brazilian Caatinga seasonally dry forests and woodlands reveals high species richness and endemism. *Journal of Arid Environments*. 174: 104079. - doi: [10.1016/j.jaridenv.2019.104079](https://doi.org/10.1016/j.jaridenv.2019.104079)
- Fick SE, Hijmans RJ (2017). WorldClim 2: new 1km spatial resolution climate surfaces for global land areas. *International Journal of Climatology* 37 (12): 4302-4315. - doi: [10.1002/joc.5086](https://doi.org/10.1002/joc.5086)
- ForestPlot.net (2021). Taking the pulse of Earth's tropical forests using networks of highly distributed plots. *Biological Conservation* 260: 108849. - doi: [10.1016/j.biocon.2020.108849](https://doi.org/10.1016/j.biocon.2020.108849)
- Ganem RS (2017). Caatinga: estratégias de conservação [Caatinga: conservation strategies]. Estudo Técnico, Consultoria legislativa, Brazil, pp. 105.
- Gerstner K, Moreno-Mateos D, Gurevitch J, Beckmann M, Kambach S, Jones HP, Seppelt R (2017). Will your paper be used in a meta-analysis? Make the reach of your research broader and longer lasting. *Methods in Ecology and Evolution* 8: 777-784. - doi: [10.1111/2041-210X.12758](https://doi.org/10.1111/2041-210X.12758)
- Guedes LPC, Uribe-Opazo MA, Junior PJR (2013). Influence of incorporating geometric anisotropy on the construction of thematic maps of simulated data and chemical attributes of soil. *Chilean Journal of Agricultural Research* 73: 1-10. - doi: [10.4067/S0718-58392013000400013](https://doi.org/10.4067/S0718-58392013000400013)
- Hengl T, Mendes De Jesus J, Heuvelink GBM, Ruiperez Gonzalez M, Kilibarda M, Blagotić A, Shangguan W, Wright MN, Geng X, Bauer-Marschallinger B, Guevara MA, Vargas R, MacMillan RA, Batjes NH, Leenaars JGB, Ribeiro E, Wheeler I, Mantel S, Kempen B (2017). SoilGrids250m: Global gridded soil information based on machine learning. *PLoS One*. 12: e0169748. - doi: [10.1371/journal.pone.0169748](https://doi.org/10.1371/journal.pone.0169748)
- Hijmans RJ, Elith J (2017). Species distribution modeling with R. R Cran Project, web site. [online] URL: <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=5ce78e97d61f3bf48d1848dcd039a7fb28ff0480>
- Hijmans RJ (2021). raster: geographic data analysis and modeling. R package version 3.4-13. [online] URL: <http://CRAN.R-project.org/package=raster>
- IBGE (2021). Mapeamento de recursos naturais do Brasil Escala 1:250.000. *Pedologia Versão 2021* [Mapping of natural resources in Brazil Scale 1:250.000. *Pedology Version 2021*]. Brazilian Institute for Geography and Statistics, Coordenação de Recursos Naturais e Estudos ambientais, Rio de Janeiro, RJ, Brazil. [online] URL: <http://www.ibge.gov.br/geociencias/download-s-geociencias.html>
- IPCC (2006). Guidelines for national greenhouse gas inventories: agriculture, forestry and other land use. Institute for Global Environmental Strategies 2: 2-59. [online] URL: <http://www.ipcc-nggip.iges.or.jp/public/2006gl/vol2.html>
- Journel AG, Huijbregts CJ (1978). *Mining geostatistics*. Academic Press, London, USA, pp. 600.
- Lima Júnior C, Accioly LJO, Giongo V, Lima RLFA, Sampaio EVSB, Menezes RSC (2014). Estimativa de biomassa lenhosa da caatinga com uso de equações alométricas e índices de vegetação [Estimation of woody biomass of the caatinga using allometric equations and vegetation indices]. *Scientia Forestalis* 42 (102): 289-298. [online] URL: <http://www.ipef.br/publicacoes/scientia/nr102/cap13.pdf>
- Lopez-Gonzalez G, Lewis SL, Burkitt M, Phillips OL (2011). ForestPlots.net: a web application and research tool to manage and analyse tropical forest plot data. *Journal of Vegetation Science* 22 (4): 610-613. - doi: [10.1111/j.1654-1103.2011.01312.x](https://doi.org/10.1111/j.1654-1103.2011.01312.x)
- Marques FA, Nascimento AF, Araújo Filho JC, Silva AB (2014). Solos do Nordeste [Northeast soils]. Embrapa Solos, Recife, PE, Brazil, pp. 8.
- Matheron G (1963). Principles of geostatistics. *Economic Geology* 58: 1246-1266. - doi: [10.2113/gsecongeo.58.8.1246](https://doi.org/10.2113/gsecongeo.58.8.1246)
- Mello CR, Viola MR, Beskow S, Norton LD (2013). Multivariate models for annual rainfall erosivity in Brazil. *Geoderma* 202-203: 88-102. - doi: [10.1016/j.geoderma.2013.03.009](https://doi.org/10.1016/j.geoderma.2013.03.009)
- Mello YR, Oliveira TMN (2016). Análise estatística e geoestatística da precipitação média para o município de Joinville (SC) [Statistical and geostatistical analysis of average precipitation for the municipality of Joinville (SC)]. *Revista Brasileira de Meteorologia* 31 (2): 229-239. - doi: [10.1590/0102-778631220150040](https://doi.org/10.1590/0102-778631220150040)
- Moonlight PW, Banda-RK, Phillips OL, Dexter KG, Pennington RT, Baker TR, De Lima CH, Fajardo L, González-MR, Linares-Palomino R, Lloyd J, Nascimento M, Prado RC, Quintana C, Riina R, Rodríguez MGM, Villela MD, Aquino ACMM, Arroyo L, Bezerra C, Tadeu Brunello A, Brien R, Cardoso D, Chao K-J, Cotta Coutinho IA, Cunha J, Domingues T, Do Espírito Santo MM, Feldpausch TR, Ferreira Fernandes M, Goodwin ZA, Jiménez EM, Levesley A, Lopez-Toledo L, Marimon B, Miatto K, Mizushima M, Monteagudo A, Soelma Beserra De Moura M, Murakami A, Neves D, Nicora Chequín R, César De Sousa Oliveira T, Almeida De Oliveira E, De Queiroz L, Pilon A, Marques Ramos D, Reynel C, Rodrigues PMS, Santos R, Särkinen T, Fernando Da Silva V, Souza RMS, Vasquez R, Veenendaal E (2021). Expanding tropical forest monitoring into dry forests: the DRYFLOR protocol for permanent plots. *Plants, People, Planet* 3: 295-300. - doi: [10.1002/ppp3.10112](https://doi.org/10.1002/ppp3.10112)
- Morais VA, Mello JM, Mello CR, Silva CA, Scolforo JRS (2017). Spatial distribution of the litter carbon stock in the Cerrado biome in Minas Gerais state, Brazil. *Ciência e Agrotecnologia* 41 (5): 580-589. - doi: [10.1590/1413-70542017415006917](https://doi.org/10.1590/1413-70542017415006917)
- Oliveira EVS (2016). Dinâmica temporal e aspectos da vegetação em uma comunidade de Caatinga [Temporal dynamics and vegetation aspects in a Caatinga community]. MSc Thesis, Universidade Federal de Sergipe, São Cristóvão, SE, Brazil, pp. 98.
- Oliveira CP, Ferreira RLC, Silva JAA, Lima RB, Silva EA, Da Silva AF, Lucena J, Santos NAT, Lopes IJC, Pessoa MML, Melo CLSMS (2021). Modeling and spatialization of biomass and carbon stock using LiDAR metrics in tropical dry forest, Brazil. *Forests* 12 (4): 1-17. - doi: [10.3390/f12040473](https://doi.org/10.3390/f12040473)
- Poorter L, Bongers F, Aide TM, Zambrano AMA, Balvanera P, Becknell JM, Boukili V, Brancalion PHS, Broadbent EN, Chazdon RL, Craven D, Almeida-Cortez JS, Cabral GAL, Jong BJJ, Denslow JS, Dent DH, Dewalt SJ, Dupuy JM, Duran SM, Espírito-Santo MM, Frandino MC, César RG, Hall JS, Hernandez-Stefanoni JL, Jakovac CC, Junqueira AB, Kennard D, Letcher SG, Licona J, Lohbeck M, Marin-Spiotta E, Martinez-Ramos M, Massoca P, Meave JA, Mesquita R, Mora F, Muñoz R, Muscarella R, Nunes YRF, Ochoa-Ganoa S, Oliveira AA, Orihuela-Belmonte E, Peña-Carlos M, Pérez-García EA, Piotta D, Powers JS, Rodríguez-Velázquez J, Romero-Pérez E, Ruiz J, Saldarriaga JG, Sanchez-Azofeifa A, Schwartz NB, Steininger MK, Swenson NG, Toledo M, Uriarte M, Bruegel M, Wal H, Veloso MDM, Vester HFM, Vicentini A, Vieira ICG, Bentos TV, Williamson GB, Rozendaal DMA (2016). Biomass resilience of neotropical secondary forests. *Nature* 530: 211-214. - doi: [10.1038/nature16512](https://doi.org/10.1038/nature16512)
- R Core Team (2021). R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. [online] URL: <http://www.R-project.org/>
- Reis AA, Diniz JMFS, Arcebi Júnior FW, Mello JM,

- Batista APB, Ferraz Filho AC (2020). Modeling the spatial distribution of wood volume a Cerrado Stricto Sensu remnant in Minas Gerais state, Brazil. *Scientia Forestalis* 48 (125): 1-13. - doi: [10.18671/scifor.v48n125.15](https://doi.org/10.18671/scifor.v48n125.15)
- Ribeiro Junior PJ, Diggle PJ, Schlather M, Bivand R, Ripley B (2020). *geoR: analysis of geostatistical data*. R package version 1:8-1. [online] URL: <http://CRAN.R-project.org/package=geoR>
- Saatchi SS, Harris NL, Brown S, Lefsky M, Mitchard ETA, Salas W, Zutta BR, Buermann W, Lewis SL, Hagen S, Petrova S, White L, Silman M, Morel A (2011). Benchmark map of forest carbon stocks in tropical regions across three continents. *Proceedings of the National Academy of Sciences USA* 108: 9899-9904. - doi: [10.1073/pnas.1019576108](https://doi.org/10.1073/pnas.1019576108)
- Sampaio EVSB, Silva GC (2005). Biomass equations for Brazilian semiarid Caatinga plants. *Acta Botanica Brasílica* 19 (4): 935-943. - doi: [10.1590/S0102-33062005000400028](https://doi.org/10.1590/S0102-33062005000400028)
- Santos RC, Castro RVO, Carneiro ACO, Castro AFNM, Pimenta AS, Pinto EM, Marinho IV (2016). Estoques de volume, biomassa e carbono na madeira de espécies da Caatinga em Caicó, RN [Volume, biomass and carbon stocks in wood from Caatinga species in Caicó, RN]. *Pesquisa Florestal Brasileira* 36 (85): 1-7. - doi: [10.4336/2016.pfb.36.85.772](https://doi.org/10.4336/2016.pfb.36.85.772)
- Scolforo HF, Scolforo JRS, Mello JM, Mello CR, Morais VA (2016). Spatial interpolators for improving the mapping of carbon stock of the arboreal vegetation in Brazilian biomes of Atlantic forest and Savanna. *Forest Ecology and Management* 376: 24-35. - doi: [10.1016/j.foreco.2016.05.047](https://doi.org/10.1016/j.foreco.2016.05.047)
- Silveira EMO, Santo FDE, Wulder MA, Arcebi Júnior FW, Carvalho MC, Mello CR, Mello JM, Shimabukuro YE, Terra MCNS, Carvalho LMT, Scolforo JRS (2019). Pre-stratified modelling plus residuals kriging reduces the uncertainty of aboveground biomass estimation and spatial distribution in heterogeneous savannas and forest environments. *Forest Ecology and Management* 445: 96-109. - doi: [10.1016/j.foreco.2019.05.016](https://doi.org/10.1016/j.foreco.2019.05.016)
- Slik JWF, Raes N, Aiba S-I, Brearley FQ, Cannon CH, Meijaard E, Nagamasu H, Nilus R, Paoli G, Poulsen AD, Sheil D, Suzuki E, Van Valkenburg JLCH, Webb CO, Wilkie P, Wulffraat S (2009). Environmental correlates for tropical tree diversity and distribution patterns in Borneo. *Diversity and Distributions* 15: 523-532. - doi: [10.1111/j.1472-4642.2009.00557.x](https://doi.org/10.1111/j.1472-4642.2009.00557.x)
- Souza DG, Sfair JC, Paula AS, Barros MF, Rito KF, Tabarelli M (2019). Multiple drivers of aboveground biomass in a human-modified landscape of the Caatinga dry forest. *Forest Ecology and Management* 435: 57-65. - doi: [10.1016/j.foreco.2018.12.042](https://doi.org/10.1016/j.foreco.2018.12.042)
- Vieira RMSP, Tomasella J, Barbosa AA, Martins MA, Rodriguez DA, Rezende FSD, Carriello F, Santana MDO (2020). Desertification risk assessment in Northeast Brazil: current trends and future scenarios. *Land Degradation and Development* 32 (1): 224-240. - doi: [10.1002/ldr.3681](https://doi.org/10.1002/ldr.3681)
- Virgens AP, Barreto-Garcia PAB, Paula A, Carvalho FF, Aragão MA, Monroe PHM (2016). Biomassa de espécies florestais em área de Caatinga arbórea. [Biomass of forest species in an arboreal Caatinga area]. *Pesquisa Florestal Brasileira* 37 (92): 555-561. - doi: [10.4336/2017.pfb.37.92.1465](https://doi.org/10.4336/2017.pfb.37.92.1465)
- Xue J, Ge Y, Ren H (2017). Spatial upscaling of green aboveground biomass derived from MODIS-based NDVI in arid and semiarid grasslands. *Advances in Space Research* 60 (9): 2001-2008. - doi: [10.1016/j.asr.2017.07.016](https://doi.org/10.1016/j.asr.2017.07.016)

Supplementary Material

Fig. S1 - Graphical distribution of regression residuals (a); and observed versus estimated biomass from the adjusted model for the Caatinga, Bahia (b).

Fig. S2 - Fitting of the theoretical semivariogram models for the residuals of the regression model, using the Maximum Likelihood method.

Link: Viana_4104@suppl001.pdf